
Electronic Data Processing (EDP) Series

Technical Conventions for the 1970 Census Summary Tapes

Introduction

Applicability of conventions. The conventions or standards described apply to data on industry compatible magnetic tapes to be made publicly available for the six "Count" summarizations and the geocode lists of 1970 census data.

Method of establishment. Conventions applicable to summary tapes prepared from Dress Rehearsal censuses were established in 1968. Prior to the establishment of those conventions, there were a series of telephone contacts with representatives of computer manufacturers in an attempt to identify the existence of some standards. The only standard that a majority of computer systems share is the FORTRAN formatted "READ" statement for handling up to 132 characters per physical record in one "READ" execution.

As specifications for the contents of various summary files became firm, the need for modification to the conventions used for Dress Rehearsal summary tapes became apparent. The modified conventions are contained in this document. The development of geocode files has generally followed the same conventions.

Conventions for Data Records

Block and sub-block. The creation of data blocks (physical records on tape) that can be handled in most computer systems by a FORTRAN formatted "READ" statement is prohibitively costly to the Bureau in producing a number of different summary tape files for the entire United States. Therefore, blocks will be of a larger size which can be divided into sub-blocks by the user with relative ease. The sub-block, which has been uniformly defined at 120 characters, is a convenient size for listing the tapes on line printers. The sub-blocks are not true logical records within a physical record because the content of the sub-blocks does not have a repetitive pattern.

The large block-size of the physical record must be small enough to be handled conveniently in a majority of computers. One size for all summary tape files proves wasteful, however, since certain files would require an excessive amount of padding. Therefore, block-size for each summary file has been fixed for the entire file at one block-size which is an even multiple of 120 characters. The minimum block-size for any file will be 720 characters; the maximum will be 2040 characters. Blocks have been so constructed that they can be sub-divided into sub-blocks of 120 characters each. Therefore, by padding, no item will be split across 120-character sub-blocks or across physical records.

Logical record. A full logical record consists of the geographic identification and all data fields for that geographic area. The logical record may consist of multiple physical records of the size specified for the particular count file. For example, a First Count logical record consists of all the geographic codes necessary to identify an enumeration district, or other area, plus the 409 data items for the area. Therefore, this logical record with padding consists of two physical records of 1800 characters each, or 30 sub-records of 120 characters each.

In the Fourth and Sixth Counts, a tape file will contain logical records of different sizes; however, the physical record size will remain the same. For example, there may be five physical records for data and two physical records for allocation counts in one logical record for one geographic area.

Fields.

Geographic identification. The geographic codes which identify a logical record will be contained in the first 120 characters of the first physical record of the logical record. These identification fields will not be repeated in subsequent physical records of the logical record. For positive identification of the beginning of a logical record, a unique character, the dollar sign (\$), will be recorded in the 120th position of the first physical record (last character position of first sub-block).

Unused positions within the geographic identification for the particular level of summarization will be padded with blanks which will produce the character for blank in any language to which it is translated.

Data fields. All data fields for a given file will be a uniform size. For any specific file this size may be six characters or eight characters. The size selected will be adequate for the file for all fields except those for aggregate value or rent. Since rounding may be an unacceptable procedure, these items are contained in fields of 12 or 16 characters, depending upon the basic choice for the file. Data fields will contain integers only. Medians or percentages will have only an implied decimal. The character set is limited to the 10 digits, minus sign, and blank.

Padding. The logical record will rarely end coincident with the end of a physical record. Therefore, the last physical record will be padded with blanks which will produce the character for blank in any language to which it is translated. It may also be necessary to pad at the end of a physical record or 120-character sub-record in order to prevent splitting of double-size data fields.

Padding will also be used within the geographic identification as described previously.

Disclosure suppression indication. In order to comply with the Bureau's legal and historical standards of data confidentiality, certain distributions, or parts thereof, will be suppressed on the 1970 summary tapes. Other distributions or single items will never be suppressed. These latter will mainly consist of simple population and housing counts.

Types of suppression on tape. When suppression occurs, it will be one of two types. (See Example 1.) With type one, the suppression code will appear in the first data item of a tabulation and all further items will be blank. With type two, the suppression code will appear in the first item, a total count of the universe suppressed will appear in the second item, and all further items will be blank. The minus sign of the suppression code is the left character in the item fields as shown in the following example.

Example 1. Suppression Codes on Summary Tapes

Data Field Size in Characters	Type of Suppression	Possible Codes in First Item in Tabulation or Sub-Tabulation	Second Item in Tabulation or Sub-Tabulation	Remainder of Tabulation or Sub-Tabulation, if any
6	1	-00001	Blanks	Blanks
6	2	-00002	Total (xxxxxx)	Blanks
8	1	-0000001	Blanks	Blanks
8	2	-0000002	Total (xxxxxxxx)	Blanks
12	1	-00000000001	Blanks	Blanks
12	2	-00000000002	Total (xxxxxxxxxxxx)	Blanks
16	1	-000000000000001	Blanks	Blanks
16	2	-000000000000002	Total (xxxxxxxxxxxxxxxx)	Blanks

Suppression indicators on the documentation. The documentation for each count will carry, next to the number of items contained in each tabulation, the indicator defining where suppression codes may appear. Either type one or type two code may appear. The possible positions for the suppression codes on the tape are explained through the following list with reference to Example 2. The x in Example 2 under "No. of Data Items" would be replaced by one of the following suppression indicators:

Suppression
Indicator

Explanation

- SX If this indicator appeared in the documentation of a tabulation, none of its data items would ever be suppressed. If the table in Example 2 had this indicator, all 210 items would always have numeric totals for the defined groupings.
- SA If this indicator appeared, each of the data items in the tabulation would have to be checked for possible replacement by the negative numeric suppression code, -0 ... 01. (If suppression occurred, it would only be type one.) If Example 2 had an SA indicator, each of the 210 data items would have to be checked.
- SO If this indicator appeared, only the first item in the tabulation would have to be checked for suppression codes. If the first item has the -0 ... 01 code, the rest of the tabulation will be blanks. A -0...02 code specifies that the second item will contain the sum total of the suppressed tabulation and all other items will be blank. If Example 2 had an SO code, the first item, e.g., Total Males 14-19 Years Now Married, would be the only item with a possible suppression code.
- S1 If this indicator appeared in the tabulation documentation, each initial item in the primary stratifications would have to be checked for a suppression code. If present, all of the data for that primary-level distribution would be suppressed independently of the other distributions. In Example 2, there are three primary stratifications: Total, White, and Negro. The three items which may contain suppression codes are item 1 (Total Males 14-19 Years Now Married), item 71 (White Males 14-19 Years Now Married), and item 141 (Negro Males 14-19 Years Now Married).
- S2 If this indicator appeared for the table, all initial items in primary and secondary stratifications need to be checked for possible suppression codes. Only those items following in the level of stratification which starts with a suppression code will be suppressed. With reference to Example 2, if this code appeared, the following six fields would have to be checked for suppression indicators:
- Total Males 14-19 Years Now Married
Total Females 14-19 Years Now Married
White Males 14-19 Years Now Married
White Females 14-19 Years Now Married
Negro Males 14-19 Years Now Married
Negro Females 14-19 Years Now Married

The rest of the items in a suppressed distribution will be blanked.

Example 2. Identification of Levels of Stratification

Example of a stratified tabulation as shown in the summary tape documentation illustrating four levels of stratification: Total, White, and Negro; Male and Female; 14-19 years, 20-24, etc.; Now Married, Widowed, etc.

Population 14 Years Old and Over by Marital Status, Age, Race, and Sex

No. of
Data Items

210 (x)¹

Count of Persons 14 and Over

By: Race (3) By: Sex (2) By: Age (7) By: Marital Status (5).

Total:

Male:

14-19 Years:

Now Married (excludes separated)

Widowed

Divorced

Separated

Never Married

20-24:

Repeat Marital Status (5)

25-34:

Repeat Marital Status (5)

35-44:

Repeat Marital Status (5)

45-54:

Repeat Marital Status (5)

55-64:

Repeat Marital Status (5)

65 years and over:

Repeat Marital Status (5)

Female:

Same as Male (35)

White:

Same as Total (70)

Negro:

Same as Total (70)

¹Location of suppression indicator. See *Suppression indicators on the documentation.*

Conventions for Header and Trailer Labels on Magnetic Tape

Header and trailer labels will conform to the "Proposed USA Standard, MAGNETIC TAPE LABELS FOR INFORMATION INTERCHANGE" printed and circulated by the United States of America Standards Institute on July 18, 1968. It should be noted that the proposed standard is similar to the IBM "standard" tape label as constituted at the time of preparation of this document.

Labels are classified into two required types. The required label types are "Volume" labels, where "Volume" coincides with a physical tape reel, and "File" labels, where a "File" has the normal meaning of a collection of related data records, more than or less than one reel.

A label is always an 80-character physical record. The following chart indicates the four-character identifiers for each type and class of label:

Type Label	Required	Optional continuation
Volume (reel):		
header	VOL1	(none)
trailer	EOV1	EOV2 thru 9
File:		
header	HDR1	HDR2 thru 9
trailer	EOF1	EOF2 thru 9

In the convention for 1970 census summary tapes, labels will be restricted as follows.

Header Labels. One Volume Header (VOL1) at the beginning of each physical tape reel following the load point.

Two File Headers (HDR1, 2) at the beginning of each file, and at the beginning of each subsequent volume (reel) of a file.

Trailer labels. Two Volume Trailers (EOV1, 2) if the reel is not the last in a multi-reel file, or is not the only reel in a file.

Two File Trailers (EOF1, 2) at the end of the file. If the reel is the only reel in a file, or the last reel in a file, only File Trailers are used.

Groups of labels. The sequence of groups of labels is as follows:

1st of 2 tapes, one file:

VOL1 HDR1 HDR2 *----data----* EOVL EOVL**

2nd of 2 tapes, one file:

VOL1 HDR1 HDR2 *----data----* EOF1 EOF2**

* = Tape Mark

Tape mark. The rules for tape mark are:

- a. One tape mark separates labels and data, occurring after all labels which precede data, and before all labels which follow data.
- b. Two tape marks appear at the end of tape after EOVL or EOF records.
- c. If a file ends within a volume and another file is to be written on that volume, then one tape mark both precedes and follows the end of file labels; the next file headers appear and are separated from their data by a tape mark.

Label content. In the system for producing these label groups on 1970 census summary tapes, a majority of the content of the labels will be constant. It is assumed that labels will be created and stored on magnetic tape and that they can be selected from this source during creation of a tape for sale. Only the variable content will be inserted by parameter card or console type-in such as external reel number, summary tape code, and 1960 State code (the File Identifier). The section on file label formats identifies the variable fields for 1970 census summary tapes.

**Technical Characteristics of Magnetic Tape Recording for
1970 Census Summary Tapes**

Characteristics	Industry Compatible Tape	
	7-Track Tape	9-Track Tape
Width	.5 inch	.5 inch
Reel Size	10.5 inch diameter; maximum 2400 foot length	10.5 inch diameter maximum 2400 foot length
Recording Mode	NRZI (nonreturn to zero)	NRZI (nonreturn to zero)
Number of Recording Tracks	7 (6 data, 1 parity)	9 (8 data, 1 parity)
Density (characters or bytes per inch)	556/800	800
Language	Binary Coded Decimal (BCD)	Extended Binary Coded Decimal Interchange Code (EBCDIC) USA Standard Code for Information Interchange (USASCII or ASCII)
Error Controls	Character - even parity Track - longitudinal redundancy check	Character (byte) - odd parity Block-Cyclical code check based on an 8-bit character which is computed from data characters during write operation and recorded at end of tape block as hardware function
Recorded Tape Representation	One 6-bit character per tape frame	One 8-bit character (byte) per tape frame
Interblock Spacing	.75 inch	.6 inch (nominal)

File Label Formats

The following material describes both the content of the file labels and the sequence of labels and data on the tapes.

Volume Header Label

<u>Field</u>	<u>Name</u>	<u>Field Length in Characters</u>	<u>Description</u>
1	Label Identifier	3	"VOL"
2	Label Number	1	"1"
3 ¹	Tape Physical Number	6	Six alphanumeric characters assigned to identify the physical reel of tape; should agree with external visual identification. Must be 6 characters other than blank.
4	Accessibility	1	Space: unlimited access.
5	Reserved for Operating System	20	System creating tape may use this space.
6	Reserved for Future	6	Spaces
7	Owner Identification	14	Spaces
8	Reserved for Future	28	Spaces
9	Label Standard Level	1	Space: label and data format on this volume require the agreement of interchange parties.

-Inter-record gap-

¹Field content is variable and inserted by operator at the time of creation of the tape file.

First File Header Label

<u>Field</u>	<u>Name</u>	<u>Field Length in Characters</u>	<u>Description</u>
1	Label Identifier	3	"HDR"
2	Label Number	1	"1"
3 ¹	File Identifier	17	Alphanumeric unique name for file. This field is recorded as follows: eight characters, the first of which must be a letter, with the remaining seven either letters or numbers followed by nine spaces. Example: ST35X1ST XXXXXXXXXXXX
4	Set Identification	6	Alphanumeric unique name for set of files of which this file is part. Must be the same in all files of a multi-file set. Must be present for single file set. Example: CT1A70.
5	File Section Number	4	Must be numeric. On the first volume a file appears on, this will be "0001". This will be augmented by 1 for each additional volume on which a file appears.
6	File Sequence Number	4	Numeric. Denotes the sequence (i.e., 0001, 0002, etc.) of files within the volume or set of volumes. In all the labels of this type for a given file, this field will contain the same number.
7	Generation Number	4	Spaces
8	Generation Version Number	2	Spaces

First File Header Label--Continued

<u>Field</u>	<u>Name</u>	<u>Field Length in Characters</u>	<u>Description</u>
9	Creation Date	6	A space, followed by two numeric characters for the year, followed by three numeric characters for the day within the year. Example: 068229 = August 16, 1968 (August 16 is 229th day of year 1968).
10	Expiration Date	6	Same format as Field 9. Will be entered as "099365".
11	Accessibility	1	Space: unlimited access.
12	Block Count	6	"000000"
13	System Code	13	Alphanumeric identification for the operating system that produces the file, supplied by Census in the following format: "CENSUS x C1 Uy" where x = machine identification (A,B,C, or D) y = tape unit identification (0, or 1)
14	Reserved for Future Standardization	7	Spaces

-Inter-record gap-

¹Field content is variable and inserted by operator at the time of creation of the tape file.

Second File Header Label

<u>Field</u>	<u>Name</u>	<u>Field Length in Characters</u>	<u>Description</u>
1	Label Identifier	3	"HDR"
2	Label Number	1	"2"

Second File Header Label-Continued

<u>Field</u>	<u>Name</u>	<u>Field Length in Characters</u>	<u>Description</u>
3	Record Format	1	"F": fixed length.
4	Block Length (Physical record)	5	"01800" (example only; some multiple of 120 between 720 and 2040).
5	Record Length ²	5	"01800" (example only; some multiple of 120, no larger than the block length in Field 4).
6	Recording Density	1	"1": 556 "2": 800
7	File Position	1	Identifies condition that caused creation of label. "0": if HDR and OPEN; if trailer and CLOSE. "1": if created because of EOJ.
8	Job/Step ID	17	Spaces
9	Tape Recording	2	"ET": even parity (7-track tape only). "BB": for 9-track tape.
10	Printer Control	1	Space: no control characters.
11	Reserved for Operating System	43	Spaces

- Tape Mark -

Data

- Tape Mark -

(Data and tape mark are followed by either the two End-of-File labels or the two End-of-Volume labels shown here. Also see section on *Groups of labels*.)

²"Record length" refers to logical record length, unless the logical records are longer than blocks (physical records). In the latter case, "record length" is always the same as "block length".

First End-of-File Trailer Label

<u>Field</u>	<u>Name</u>	<u>Field Length in Characters</u>	<u>Description</u>
1	Label Identifier	3	"EOF"
2	Label Number	1	"1"
3-11	Same as corresponding fields in HDR1		
12	Block Count	6	Numeric. Number of data blocks (exclusive of labels and tape marks) since the preceding HDR label group.
13-14	Same as corresponding fields in HDR1		

-Inter-record gap-

Second End-of-File Label

<u>Field</u>	<u>Name</u>	<u>Field Length in Characters</u>	<u>Description</u>
1	Label Identifier	3	"EOF"
2	Label Number	1	"2"
3-11	Same as corresponding fields in HDR2		

First End-of-Volume Label

<u>Field</u>	<u>Name</u>	<u>Field Length in Characters</u>	<u>Description</u>
1	Label Identifier	3	"EOV"
2	Label Number	1	"1"
3-11	Same as corresponding fields in HDR1 of current file		
12	Block Count	6	Numeric. Number of data blocks since the preceding HDR label group.
13-14	Same as corresponding fields in HDR1		

-Inter-record gap-

Second End-of-Volume Label

<u>Field</u>	<u>Name</u>	<u>Field Length in Characters</u>	<u>Description</u>
1	Label Identifier	3	"EOV"
2	Label Number	1	"2"
3-11	Same as corresponding fields in HDR2		

Character Set for the 1970 Census Summary Tapes

A White House memorandum to heads of departments and agencies, signed by President Johnson on March 11, 1968, approved the United States of America Standard Code for Information Interchange (ASCII) as the Federal standard for magnetic tape. The Census Bureau sells magnetic tape files written in a subset of ASCII consisting of the ten decimal digits, the twenty-six alphabetic symbols, the blank or space, the Tape Mark, and six of the possible special characters. Also, since many data users have not yet been able to convert to ASCII codes because of their present hardware or software, the Census Bureau has agreed to release tapes in Binary Coded Decimal (BCD) and Extended Binary Coded Decimal Interchange Code (EBCDIC).

Conversion for Public Use. The Census Bureau processes data for internal use of UNIVAC 1107 and 1108 computers which use code schemes called Binary Coded Decimal, excess three (BCD XS-3), and FIELDATA. For compatibility with the Federal standards and industrial practice, conversion of the codes is performed before release to the public. An IBM 360, Model 40 is used to convert from BCD to ASCII and EBCDIC, and an IBM 1401 and the UNIVAC 1107 and 1108 computers convert from BCD XS-3 to BCD.

Code Representation. The attached character conversion chart contains the forty-three character graphics and their coded representation in ASCII, BCD, and EBCDIC. The six-bit character code is shown in octal representation; the others are shown in hexadecimal.

Character Conversion Chart

Conversion Code for 43 Characters in ASCII, BCD, and EBCDIC

<u>Character graphic</u>	<u>ASCII</u>	<u>BCD¹ (octal)³</u>	<u>EBCDIC² (hexadecimal)⁴</u>
blank (space)	20	20 ⁵	40
- (minus)	2D	40	60
0	30	12	F0
1	31	01	F1
2	32	02	F2
3	33	03	F3
4	34	04	F4
5	35	05	F5
6	36	06	F6
7	37	07	F7
8	38	10	F8
9	39	11	F9
, (comma)	2C	33	6B
. (period)	2E	73	4B
A	41	61	C1
B	42	62	C2
C	43	63	C3
D	44	64	C4
E	45	65	C5
F	46	66	C6
G	47	67	C7
H	48	70	C8
I	49	71	C9
J	4A	41	D1
K	4B	42	D2
L	4C	43	D3
M	4D	44	D4
N	4E	45	D5
<u>O</u>	4F	46	D6
<u>P</u>	50	47	D7
<u>Q</u>	51	50	D8
<u>R</u>	52	51	D9

Character Conversion Chart-Continued

Conversion Code for 43 Characters in ASCII, BCD, and EBCDIC

<u>Character graphic</u>	<u>ASCII</u>	<u>BCD¹ (octal)³</u>	<u>EBCDIC² (hexadecimal)⁴</u>
S	53	22	E2
T	54	23	E3
U	55	24	E4
V	56	25	E5
W	57	26	E6
X	58	27	E7
Y	59	30	E8
Z	5A	31	E9
\$ (dollar)	24	53	5B
* (asterisk)	2A	54	5C
/ (slash or virgule)	2F	21	61

¹This is a 6-bit code known as binary coded decimal, used originally in earlier model IBM computers and peripheral devices.

²This is an 8-bit code, known as extended binary coded decimal interchange code, which originated with IBM Systems 360.

³In the octal representation of each 6-bit code the three high order bits are represented in the most significant digit of the octal code, and the three low order bits are represented in the least significant digit of the octal code. The range of each octal digit is 0-7.

⁴In the hexadecimal representation of the 8-bit code, the four most significant bits are represented in the most significant position of the hexadecimal code, and the four least significant bits are represented in the least significant position of the hexadecimal code. The value range of each hexadecimal character is 0-15, where the values 10, 11 - - - 15 are defined by the letters A, B - - - F, respectively.

⁵Octal 20 for blank, in BCD, appears only on tape. Internally this code is 00 and the difference between the internal and external representations is handled automatically by the hardware. Where BCD is generated internally in the UNIVAC 1107 or 1108, and written to tape without the hardware translation, the octal 20 must be created in core.